

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Procedia - Social and Behavioral Sciences 34 (2012) 219 – 225

Procedia
Social and Behavioral Sciences

Languages, Cultures and Virtual Communities
Les Langues, les Cultures et les Communautés Virtuelles

Can simultaneous reading and listening improve speech perception and production? An examination of recent feedback on the SWANS authoring system

Anthony Stenton*

LAIRDIL, Université Toulouse I, Capitole, 2 rue du Doyen-Gabriel-Marty, 31042 Toulouse Cedex 9, France

Abstract

Recent CNRS-financed CALL research in Toulouse offers hope for improving L2 listening perception and spoken production through new on-screen, simultaneous reading and listening techniques and an emphasis on lexical stress visual annotation as a neglected key to acquiring basic spoken intelligibility. Although the research has concentrated principally on the French/English language pair its implications are potentially important for the study of all languages and the SMIL-based authoring system developed is language independent and has been used for producing prototype learning activities in German, Dutch, Spanish, French, Italian, Finnish, Russian and Arabic. The authoring system SWANS (Synchronised Web Authoring Notation System) uses metaguiding via highlighted annotated text together with synchronised sound to inhibit L1 interference during the reading process. The shared networked use of documents generated with SWANS constitutes an innovation in teacher training by allowing experts in local L1 interference problems to share their knowledge of lexical stress on a Europe-wide basis.

© 2012 Published by Elsevier B.V. Selection and/or peer-review under responsibility of EUROCALL2010 Scientific Committee. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: lexical stress patterns; synchronisation; AV perception; mother-tongue interference; SMIL; multimodal reading

1. Introduction

Alphabets are generally invented in a monolingual, monocultural context. The complex, not to say contradictory, needs of the multilingual community are ignored. Consequently, pedagogical engineers and designers of activities for foreign language learning must strive to destabilize the effects of early

* Anthony Stenton. Tel.: + 33 (0) 561 867 225
anthony.stenton@univ-tlse1.fr

conditioning – ‘fossilised’, intractable L1 sound / symbol associations – and prepare the human brain for a host of new, unfamiliar L2 cognitive activities. Some learning problems associated with technology defy ‘quick-fix’ solutions. Inertia and a certain feeling of helplessness may explain pedagogical indifference to the permanently obsolete QWERTY computer keyboard but how can language engineers remain indifferent to the greatest problem in human communication: the intelligibility of the spoken word? If the Latin alphabet is a problem for foreign language learning then it deserves to be redesigned and adapted to the needs of the multilingual community.

In the context of static, paper-based Gutenberg technology, attempting to change learning and reading by changing the colour of the alphabet, such as Caleb Gattegno’s (e.g. 1968) Silent Way method, appeared presumptuous. The technology inside the classroom was limited to wall posters and outside the classroom was simply unavailable. In the plastic, colourful, flexible, dynamic, nomadic world of the 21st century LCD screens, SMIL language (W3C, 1999), smart telephones, e-readers and other omnipresent mobile devices, vast ‘AV perception’ manipulations including animation and sound/text synchronisation become possible. In this article we chart some of the recent pedagogical progress made with visually-improved typography and sound synchronised annotations which must be seen in the wider context of developing SWANS as an automatic ‘syllabifier’ using increasingly reliable algorithms (95% accuracy) from text-to-speech programmes for the generation of annotated text showing the lexical stress patterns of European languages.

2. The SWANS authoring system (Synchronised Web Authoring Notation System)

The SWANS project united 12 researchers from the fields of linguistics, psycholinguistics, computer science, cognitive science, acoustics and human-computer interface design from 4 Toulouse laboratories (LAIRDIL, LTC, LAAS-CNRS, LORDAT) financed in the context of the CNRS-TCAN programme 2004-2006. The working hypothesis was that students must be helped to see the lexical stress patterns, including weak vowels, which they patently did not always hear through L1 interference. A number of potential visual models were examined and rejected:

Graphic models of sound waves were rejected on the grounds of excessive complexity and debatable accuracy. They display:

- Pitch curves (which measure height);
- Wave curves or oscillograms (which measure volume or intensity);
- Spectral analysis (for comparing different voice qualities).

Manifested in excellent and scientifically useful programmes such as ‘PRAAT’ or ‘Winpitch’, or reduced to the level of a sales promoting gadget in ‘Tell me more’ and others, they:

- required expert knowledge to interpret;
- were incomplete or ambiguous (displaying only one or two parameters rather than simultaneously measuring length, height and volume like the human ear);
- were excessively time consuming to generate and decipher;
- could not justify the time investment for non specialist language learners even if automatic text alignment could be implemented.

By contrast, SWANS-type textual annotations, which are relatively modest and discreet, aim above all to respect the essential requirements of a low learning curve for teachers and students and the possibility of rapid retention of stress patterns via dual coding to accelerate learning. According to the theories of

Paivio (1991), Sweller (1999) and Mayer (2001), such dual coding based on sound and typography should lead to better learning for novice learners.

WILLARD ROWLAND: Well, I think it's **always** the **problem** of that in a **profit-oriented** **environment**. There is a great need for the **producers** to **produce** not **necessarily** high **quality** **programming** but large **numbers** of **audience**. So in the **newspaper** it's a **question** of the **circulation** size and in the **broadcasting station** it's a **question** of the **ratings**. And to the **extent** that the **companies** are more and more **concentrated**, the **pressures** for that sort of **production** of **profit** are **higher** and **higher**. It's **exacerbated** as well in many **countries**- **particularly** in the **US** but I think also to a certain **extent** now in **Europe** as well.

Figure 1. Textual annotation in SWANS

Key: blue for primary stress (size 24); orange for reduced vowels (size 12); purple for secondary stress (size 22). Ordinary text is black (size 18).

The SWANS authoring system uses typographical techniques, changes of letter size and colour, and line-by-line synchronisation, to help students notice and memorise the target lexical stress patterns. We refer the reader to (Stenton *et al.*, 2005a) for an account of crucial design choices in terms of text size and colour, and to (Stenton *et al.*, 2005b) for an account of technical considerations in the creation of sound synchronised web pages with the authoring system. Throughout its development stages SWANS has been tested by teachers across Europe in CercleS language centres and we are grateful for feedback on multilingual interference questions and the notion of training teachers from different countries through shared networked annotation of the same basic documents.

The invention of SMIL language (W3C, 1999, 'Synchronised Multimedia Integration Language') which permits fine-tuned synchronisation within a web page (measured for the first time in hundredths of a second), constitutes an important breakthrough.

On the computer or LCD screen different techniques of synchronisation have emerged in software in the last 12 years:

1. Subtitles where a short line of text appears and disappears fleetingly in time with the soundtrack.
2. Karaoke, which is usually based on word by word synchronisation and which is visually distracting through the jerking movement of the animation.
3. Highlighting line-by-line, which uses a band of colour behind the particular words which the listener is listening to.

For the SWANS team, synchronisation through line-by-line highlighting of annotated text offered potential advantages which few language teachers have considered. When silently reading an L2 text, all language students subvocalise to a certain degree, that is to say they 'hear' much of the text in their head. As this process engenders L1 interference, what students hear undoubtedly suffers from perceptual distortion. When L2 scripts and L2 soundtracks are synchronised and annotated, however, the reading brain can no longer subvocalise in the same uninhibited way. The L2 soundtrack and the textual annotation interfere with the L1 interference which obliges the brain to choose between perceptually distorted subvocalisation and the L2 sounds the ear is busy decoding. Simultaneous listening and reading

is not just multitasking, it is creating a destabilising dilemma for the brain which is receiving mixed messages from the eyes and ears. Whether or not this destabilising experience can become a positive reinforcement of the authentic L2 speech patterns, accelerate the elimination of L1 interference fossils such as *development* and thereby enhance intelligibility in oral production, is precisely the challenge our research set out to measure.

Line by line synchronisation of text through highlighting also offers a degree of ocular comfort which should not be underestimated. Synchronisation, we suggest, represents a lightening of the cerebral burden as the eyes are *guided* and can focalise more easily on essential linguistic clues. Feedback from students specifically refers to reduced eye fatigue.

3. Testing synchronisation and annotation techniques

Testing the dual coding effect of synchronisation and SWANS annotation techniques, included formal testing of perception and oral production before and after training, and questionnaires to obtain student feedback.

Early testing of SWANS took place with over 250 students (CEFR levels A2 to B2) from three universities and a secondary school between 2004-2005. Test results are analysed in (Stenton et al, 2005a), and (Stenton et al. 2005b). More recent testing (2007-2010) concerns over 300 students in postgraduate Master courses (levels B1 to C2). Student L1s were approximately 80% French. 20 % of students were of some 30 different nationalities: European, North African, African, Chinese.

4. Pedagogical methods

10-week courses using SWANS documents aimed to increase awareness of English lexical stress patterns. Video film of fluent English-speaking European politicians and managers who ignored English stress patterns were used to demonstrate problems.

Pedagogical exploitation was organised around 5 activities:

1. Reading out loud from the unannotated script in pairs, with mutual correction.
2. Annotating scripts on paper after listening to the sound, with self-correction.
3. Carrousel activities: repeated 3-minute oral summaries on research topics in constantly changing pairs. The same presentation is made 4 times to different partners. After the first presentation the *student stops worrying about content which frees brain resources for concentrating on spoken form.*
4. Distance teacher correction of student annotated keywords, followed up by 'Powerpoint' oral presentations in class.
5. EXPLICS Internet case studies where oral performance was continually monitored (Website managed by CercleS in Goettingen).

By moving from formal CALL-assisted academic analysis of stress patterns to purposeful, voluntary student organised presentations and case studies, we attempted to construct a context where students would become aware of a need for ear training, of the need to filter out the 'fossils' in their oral production whose credibility could easily be undermined by an accumulation of misplaced lexical stress.

5. Testing procedures and student performance

After this initial training, the following profiles emerged. Tests of perception and oral production revealed striking parallels. If scores were high in perception they were almost always high in oral

production too. By contrast annotation tests, where students were requested to manually annotate a text by underlining stressed syllables and reduced vowels, were far less predictable. Annotation tests included items which were unsegmented (*fantastic*) and pre-segmented (*fan-tas-tic*). Tests of listening perception were based on multiple choice tests where the students listened to both isolated words and to contextualised items in dialogues before deciding where the lexical stress was placed. Tests of spoken English used student recordings of individual items and of phrases and 3-minute classroom oral presentations prepared in advance by the annotation of selected keywords chosen by the students and corrected via e-mail by the teacher.

What was striking, in France, during initial pre-course testing, was the omnipresent nature of L1 interference. The term ‘stress deafness’ is by no means excessive even if it does not refer to a real physical impediment. Low scores in textual annotation, listening and spoken production were common among students across the spectrum of previously tested levels, that is to say even relatively fluent students were regularly judged unintelligible owing to an accumulation of lexical stress errors.

Among annotation errors in keywords chosen by the students themselves for their presentations (level B2) we noticed (underlining shows what the student *thought* was a primary stress):

- Offence, infringement, analysis, linguistics (from a doctoral student studying ‘linguistics’) economics (from an economist), enigmatic, distressed, paranormal, understanding, incomprehensible, accident, sequence, realisation, development

From a French point of view, this unconscious attempt to reinstate lexis (originally ‘borrowed’ by the English?) into its ‘rightful’ acoustic orbit may seem very natural. Notice that many of the above annotation errors were *not* always present in the listening perception tests or in the oral production of the students in question.

High performance in all activities was relatively rare suggesting that dealing with fossilised problems of pronunciation after 8 to 10 years of deviant acoustic reinforcement probably requires more than the 10 weeks or 30 hours available.

6. Results

Globally dual coding improves performance. 75% of students exposed to dual coding over 10 weeks scored at least 10% higher than in initial lexical stress recognition tests. 25% showed no change. Student feedback showed the synchronised multimodal experience in SWANS documents was appreciated and given the choice between monomodal techniques (text and sound separated) or multimodal techniques (SWANS), students preferred to use SWANS. This was a potentially important finding as the multimodal synchronised experience is theoretically more demanding. Annotations appear to improve short-term memorisation in most cases. Perception of reduced vowels in listening tests and in written tests was often more difficult for students than the perception of stressed syllables suggesting neglect in early training.

Oral testing based on a 3-minute presentation showed significant improvement in controlled conditions for certain individual students and a small global improvement after preparation via keyword annotations. On the other hand, the results of oral testing in the context of subsequent spontaneous conversation did not confirm the idea of improved oral production. ‘Fossils’ returned to the L2 oral performance perhaps because concentrated filtering efforts were reduced and the language content moved away from the controlled area of the presentation.

7. Discussion

Sociologist looking back on the history of the early 21st century may well describe the period as one of fundamental cultural change. The written culture which has reliably produced, trained and educated political and social elites for centuries is losing credibility and ceding ground to a new, ill-defined, oral culture which draws heavily on instantly available knowledge via Internet. As the speed of technological change accelerates, teachers have to admit the limits of their knowledge for handling problems which do not yet exist or for describing solutions which have yet to be invented. In such a context, teachers place the emphasis on learning skills rather than content and on engaging student motivation around the definition of those skills. We have argued that for languages one of the key, and yet paradoxically neglected, skills lies in improving the oral performance required for this brave, new, cosmopolitan 21st century world. More accurate perception and production means improved spoken intelligibility and therefore improved communication itself. The SWANS authoring system has already demonstrated a contribution in this field by allowing novice teachers to generate appropriate documents simply and quickly without code manipulations. The on screen textual annotation of syllables attempts to achieve via simultaneous heightened stimulation of eyes and ears what (Tomatis, 1971) claimed to achieve through more cumbersome and relatively complex auditory manipulations with an 'Electronic Ear': the triggering of a mental control mechanism permitting the readjustment of voice quality (in particular with regards stressed syllables and weak vowels) and thus a clearer, more comprehensible, L2 vocal production.

8. Conclusion

Although initial testing of the SWANS authoring system permits a modest 'yes' to the question 'can dual coded texts improve perception and production?', we are convinced that the battle for the eyes and ears of 21st century students will remain an arduous one. Modifying L1 reading practice when tackling an L2 text means fighting deeply engrained habits that have taken years to put in place. The brain's resistance to novelty is not a sign of conservatism but simply a sign of natural mental health. As Ong puts it, "Freeing ourselves from typographic bias is probably more difficult than any of us can imagine" (Ong, 1982: 77). Fortunately, new technology holds out increasingly powerful and increasingly nomadic devices to face up to this very real 21st century challenge. As the publishing world prepares to swamp educational markets with e-Readers (22 million in 2012) it is time for the CALL community to lend a hand to transform such devices into genuinely useful tools for reading foreign language texts. Electronic readers for European citizens should offer automatic annotation of stressed syllables and weak vowels for texts in all European languages just as word processing programmes offer spelling correctors. The technology and know-how exist, the finances deserve to be found. The limited nature of traditional European project financing, whereby 9 million euros are divided up annually into relatively small and pedagogically unambitious projects, needs to be revised. In the late 20th century the educational world stood by passively while the IBM-led private sector took decades to put spelling correctors in place. The quality of such learning tools can always be called into question but today their fundamental utility is self-evident. Much as teachers are loathe to admit, spelling correctors have made the greatest ever contribution to correct written form. We believe that syllable annotation represents an equivalent tool for the acquisition of correct spoken forms. Paradoxically, by dint of massive investment and the profit motive we have learnt how to make machines more intelligent than our teachers, just as we taught computers to beat Garry Kasparov at chess, but there is nothing inevitable or 'natural' about such a process. The intelligence of automatic syllable recognition can be re-engineered to make sound patterns more recognizable and memorable for ordinary people, to make our students more intelligent and communicative rather than confining progress to machines. From the point of view of the human race,

there is little doubt that students represent a worthier cause. Such a process might not generate the short term profits so eagerly sought by computer manufacturers but the educational benefits would most certainly be a worthwhile, permanent and very human compensation.

References

- Exploiting Internet Case Studies and Simulation templates for language teaching and learning (EXPLICS)*. (n.d.). Retrieved from www.zess.uni-goettingen.de/explics/
- Gattegno, C. (1968). *The silent way for teaching English as a second language*. Reading: Educational Explorers.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.
- Ong, W. (1982). *Orality and literacy: The technologizing of the word*. London: Methuen.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255-287.
- Stenton, A., Péchou, A., Vaillant, C., & Tricot, A. (2005a). Effet du double codage synchrone de l'accentuation en L2 selon des modalités de restitution du sujet. *Didcog: 1e colloque international de didactique cognitive*, Toulouse, Université Mirail. Retrieved from <http://acoustic31.univ-tlse2.fr/didcog/Intro.html>
- Stenton A., Tazi, S., & Tricot, A. (2005b). Document design for learning English pronunciation: The swan-song of Gutenberg? *ED-MEDIA 05: World Conference on Educational Multimedia, Hypermedia and Telecommunication*, Montreal. Retrieved from <http://www.aace.org/conf/edmedia/default.htm>
- Sweller, J. (1999). *Instructional design in technical areas*. Camberwell: Australian Council for Educational Research.
- Tomatis, A. (1971). *L'oreille et le langage*. Paris: Point-Seuil.
- W3C. (1999.). *Synchronized multimedia integration language (SMIL 3.0)*. Retrieved from <http://www.w3.org/TR/SMIL3/>